KINECT V2 AND RGB STEREO CAMERAS INTEGRATION FOR DEPTH MAP ENHANCEMENT

R. Ravanelli[a,] [*], A. Nascetti[a], M. Crespi[a]

[a] Geodesy and Geomatics Division, DICEA - University of Rome "La Sapienza", Rome, Italy
<roberta.ravanelli, andrea.nascetti, mattia.crespi>@uniroma1.it

Commission V WG V/4

KEY WORDS: 3D modelling, real-time, range cameras, RGB stereo cameras, sensor integration

ABSTRACT:

Today range cameras are widespread low-cost sensors based on two different principles of operation: we can distinguish between Structured Light (SL) range cameras (Kinect v1, Structure Sensor, ...) and Time Of Flight (ToF) range cameras (Kinect v2, ...). Both the types are easy to use 3D scanners, able to reconstruct dense point clouds at high frame rate. However the depth maps obtained are often noisy and not enough accurate, therefore it is generally essential to improve their quality. Standard RGB cameras can be a valuable solution to solve such issue. The aim of this paper is therefore to evaluate the integration feasibility of these two different 3D modelling techniques, characterized by complementary features and based on standard low-cost sensors.
For this purpose, a 3D model of a DUPLO[TM] bricks construction was reconstructed both with the Kinect v2 range camera and by processing one stereo pair acquired with a Canon Eos 1200D DSLR camera. The scale of the photgrammetric model was retrieved from the coordinates measured by Kinect v2. The preliminary results are encouraging and show that the foreseen integration could lead to an higher metric accuracy and a major level of completeness with respect to that obtained by using only separated techniques.

1. INTRODUCTION

Nowadays 3D modelling is a subject of great interest in very different fields, from industry to robotics, from cultural heritage to medicine. Thus, it is essential to find instruments providing accurate, dense and low cost 3D data, at high frame rates and in a user-friendly way.

To satisfy these requirements, a single method is not sufficient. For this reason, in this work we present the results of a preliminary test, performed to evaluate the integration feasibility of two different 3D modelling techniques: range imaging and traditional stereo vision. In fact, the integration between these two methods can offer many advantages since they are characterized by complementary features.

On the one side, range cameras are low cost and easy to use imaging sensors, able to measure distances of the scanned scene at high frame rate. They can be used as 3D scanners to easily reconstruct dense point clouds practically in real time. However these sensors show issues with transparent and very reflective surfaces. Furthermore, the depth maps obtained are generally noisy and the finer details of the resulting 3D models are often smoothed.

On the other side, stereo vision is an established technique, but the resulting 3D models are mostly incomplete in low texture regions. In addition, the processing needs an external scale that the user must provide and it is often computationally onerous and time consuming.

Therefore the leading idea, which will be developed in future works, is that a preliminary depth map of the investigated object can be obtained in real-time through a low-cost range camera. This depth map will be employed as a coarse 3D model for classical stereo processing, which will add the details coming from the stereo images acquired through standard cameras.

[*]Corresponding author.

In details, the coarse depth map acquired by the range camera will be the geometrical constraint for the subsequent Semi Global Matching (SGM) algorithm that will compute the stereo disparity map. In this way the efficiency of the dense matching algorithm will be increased.

This paper is organized as follows. In Paragraph 2 a short review of the state-of-the art is illustrated; in Paragraph 3 the preliminary test is described and its results are discussed in Paragraph 4. Finally some conclusions are outlined.

2. RELATED WORKS

The topic of integration between products from range and RGB cameras has been investigated for several years and a substantial literature is available (Nair et al., 2013), mainly considering middle to high cost professional sensors. Our goal here is to reconsider the already obtained methodological results under the light of new available low-cost sensors that can be integrated in a flexible solution.
Hereafter we just summarize the most relevant achievements of the previous works:

- (Zhu et al., 2008) combines a professional ToF sensor with two CCD cameras, introducing a method for improving the range camera manufacturer calibration

- (Yilmaz and Karakus, 2013) proposes a system formed by a professional three stereo camera and a low-cost SL range camera and suggests an accuracy improvement of the resulting 3D model through a stereo visual odometry integration

- (Evangelidis et al., 2015) presents a high-resolution stereo matching algorithm guided by low-resolution depth data, that helps the algorithm to compensate for its difficulty in estimating disparities over weakly textured areas

3. PRESENTATION AND DISCUSSION OF THE PRELIMINARY TEST

For the purposes of this study, a 3D model of a DUPLO™ bricks construction was reconstructed both with the Kinect v2 range camera and by processing one stereo pair acquired with a Canon EoS 1200D DSLR camera. The two 3D models were then fused, obtaining the integrated model.

In particular, the Kinect v2 is a ToF range camera: it evaluates the distance from the scene by measuring the time of flight that an infrared wave takes to travel from the sensor itself to the scanned scene and back. We developed our own software tool to download the 3D data with the Microsoft SDK and the model point cloud was reconstructed from the depth map (see Figure 1) acquired in a single frame, since the final aim is the near real time integration.
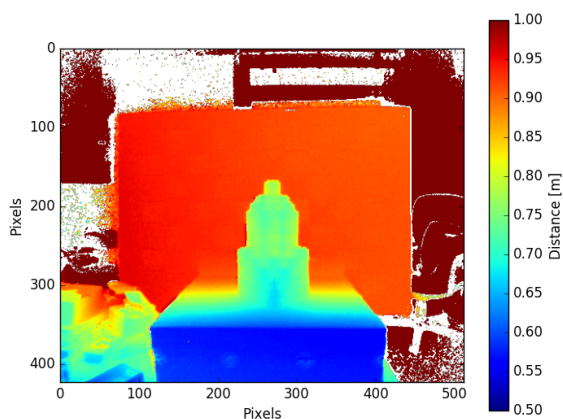

Figure 1: Depth map acquired by the Kinect v2 range camera

Regarding the stereo model, one stereo pair was acquired with a proper stereo baseline (base-to-height ratio of about 0.3) and it was processed with the Agisoft PhotoScan™ photogrammetric software. The approximated model scale was estimated by imposing the values of the coordinates measured by the Kinect v2 to eight points collimated on both images (see Figure 2).

The fusion of the Kinect v2 and the photogrammetric models was performed trough the CloudCompare (Girardeau-Montaut, 2016) 3D point cloud and mesh processing software. Since the two points clouds were already in the same reference system, the co-registration was only refined using the Iterative Closest Point (ICP) algorithm (Best and McKay, 1992), which estimated the parameters of the residual roto-translation (with scale) transformation.

4. RESULTS

To assess the metric quality of the results obtained, both the integrated model and the models reconstructed with the single 3D modelling technique were compared with the reference model of the DUPLO bricks construction.

The dimensions of bricks were measured with a vernier caliper and the reference mesh model (see Figure 3) was reconstructed with a standard CAD software.

In particular, precision and accuracy were evaluated in terms of signed distances (positive inside and negative outside the reference mesh surface) of the 3D model points from the reference mesh.

(a) First image

(b) Second image

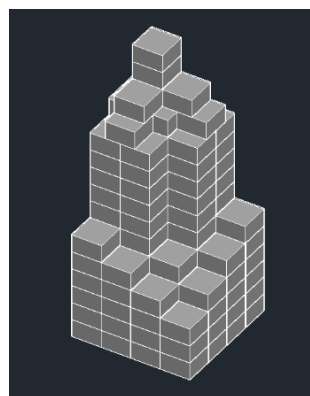Figure 2: Acquired stereo pair

Figure 3: CAD reference model

Concerning the completeness assessment, it was based on the criterion of 2D grid occupancy: the point cloud was projected onto a regular grid mesh (with a posting of 0.004 m × 0.004 m) and the completeness index was computed as the ratio of the non empty cell number to the grid total size.

Model	Distance mean [m]	Distance std [m]	Completeness [-]
Photogrammetric	0.000	0.002	38%
Kinect v2	0.004	0.015	67%
Integrated	0.001	0.003	69%

Table 1: Distance statistics

The photogrammetric model is the most accurate and precise, as the statistics of distances show (see Table 1), and it is reported in Figure 4(a). Anyway the borders between the bricks are reconstructed very well, thanks to the high texture variation, but the model is not complete in correspondence of the areas with uniform texture (single bricks).

(a) Photogrammetric model results

(b) Photogrammetric model distance histogram

(c) Kinect v2 model results

(d) Kinect v2 model distance histogram

(e) Integrated model results
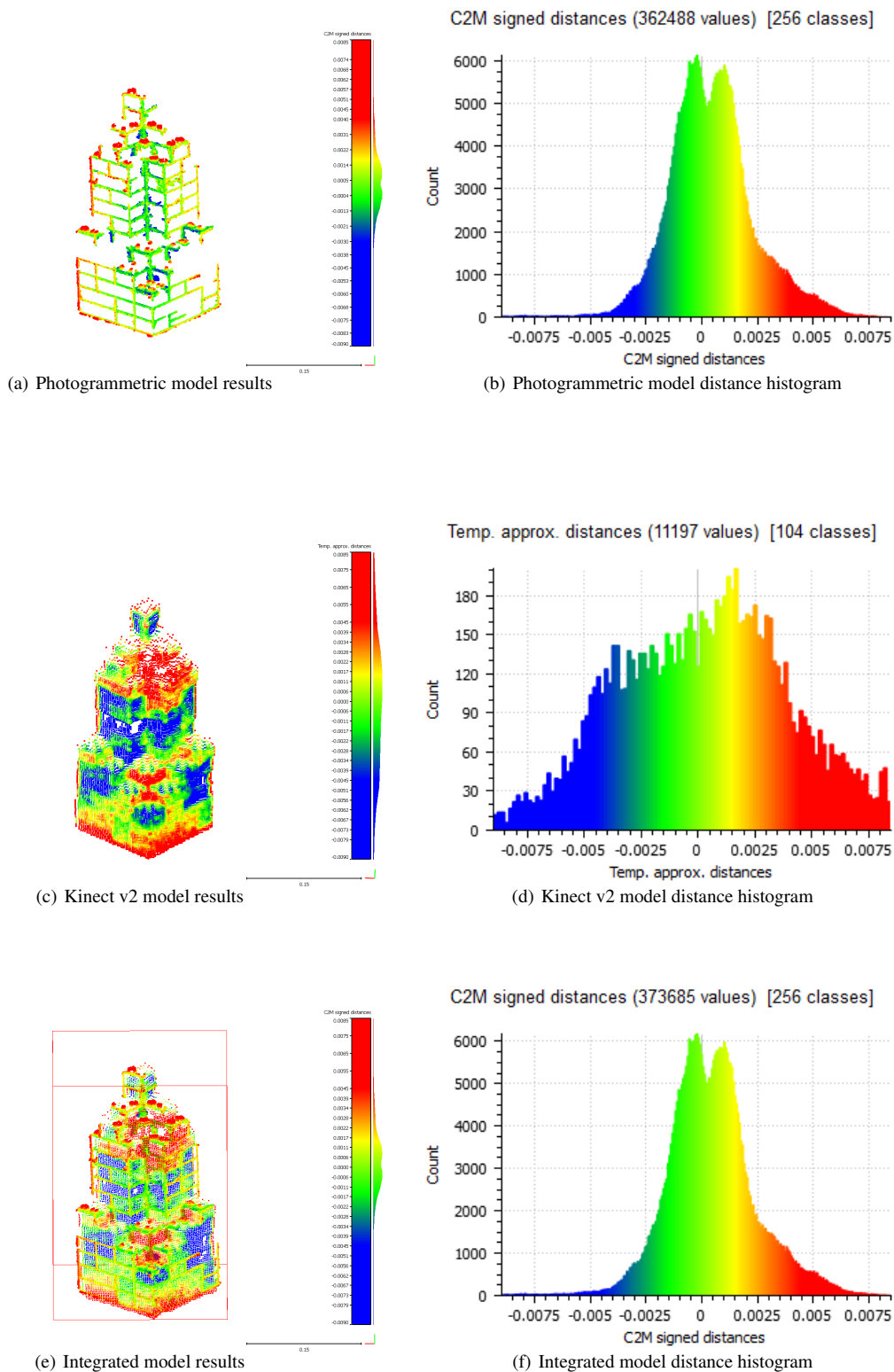
(f) Integrated model distance histogram

Figure 4: Results

As regards the model reconstructed by the Kinect v2, it is less accurate and less precise: the mean and the standard deviation of distances from the reference mesh model reach the values of 0.004 m and 0.015 m respectively. The details of the bricks are generally less recognizable (see Figure 4(c)) and the model shows some inaccuracies (flying pixels) on the edge of the DUPLO construction, where there are high depth variations.

Finally the integrated model (see Figure4(e)) preserves the accuracy and the precision of the photogrammetric model, but it also presents the greatest level of completeness, provided by the contribution of the Kinect v2 sensor.

5. CONCLUSIONS AND FURTHER DEVELOPMENTS

Our preliminary results are encouraging and show that this integrated approach leads to higher metric accuracy of the final 3D model with respect to that obtained by only using a range camera and to an higher level of completeness respect to that obtained by only processing a stereo image pair.

Future works will deeply investigate the effects of distance from the scanned object and automate the processing procedure by implementing the method previously described.

REFERENCES

Best, P. J. and McKay, N. D., 1992. A Method for Registration of 3-D Shapes. IEEE Transactions on pattern analysis and machine intelligence 14(2), pp. 239–256.

Evangelidis, G. D., Hansard, M. and Horaud, R., 2015. Fusion of Range and Stereo Data for High-Resolution Scene-Modeling. Pattern Analysis and Machine Intelligence, IEEE Transactions on 37(11), pp. 2178–2192.

Girardeau-Montaut, D., 2016. Cloud Compare3D Point Cloud and Mesh Processing Software - Version 2.7.0. Open Source Project.

Nair, R., Ruhl, K., Lenzen, F., Meister, S., Schäfer, H., Garbe, C. S., Eisemann, M., Magnor, M. and Kondermann, D., 2013. A Survey on Time-of-Flight Stereo Fusion. In: Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications, Springer, pp. 105–127.

Yilmaz, O. and Karakus, F., 2013. Stereo and kinect fusion for continuous 3d reconstruction and visual odometry. In: Electronics, Computer and Computation (ICECCO), 2013 International Conference on, IEEE, pp. 115–118.

Zhu, J., Wang, L., Yang, R. and Davis, J., 2008. Fusion of Time-of-Flight Depth and Stereo for High Accuracy Depth Maps. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, IEEE, pp. 1–8.